UNITED STATES DEPARTMENT OF COMMERCE
Bureau of the Census
Washington, DC 20233-0001

MASTER FILE

May 3, 1999

DSSD CENSUS 2000 PROCEDURES AND OPERATIONS MEMORANDUM SERIES R.-8

MEMORANDUM FOR	Robert W. Marx Chief, Geography Division
From:	Howard Hogan Chief, Decennial Statistical Studies Division
Prepared by:	Peter P. Davis and Thomas Mule Sample Design Team, Decennial Statistical Studies Division
Subject:	Census 2000 Specifications for Block Cluster Formation–Reissue

This memorandum is being reissued for inclusion in the official *DSSD Census 2000 Procedures and Operations Memorandum Series*. All text in the version of this memorandum distributed on February 16, 1999 remains the same in this version.

I Introduction

As a preliminary stage of the Post-Enumeration Survey (PES) design, this memorandum addresses the formation of block clusters. The PES program for Census 2000 requires that a sample of housing units be selected for intensive reinterviews. The goal of block clustering is to form a group of blocks that average 30 housing units and can be identified by interviewers in the field. PES design consists of dividing the United States into block clusters, groups of geographically contiguous blocks and housing units. Then a sample of block clusters will be drawn. Geography Division (GEO) will perform the clustering. Section II lists the assumptions utilized in the block clustering design. Section III lists the input files/systems that will be used in clustering. Section IV describes the specifications for the formation of block clusters for Census 2000. Section V describes certain identification processes in the block clustering procedure. Section VI lists the desired outputs needed for clustering verification and monitoring future sampling operations. Section VII involves the after sampling file preparation.

These specifications should be used to flowchart the process, to generate further discussion on requirements, to identify and finalize the record layouts of input and output files, and to write computer software to implement the methodology. During and after a testing phase, it is likely that changes to the specifications will be necessary.

Please direct any questions about these requirements to
Peter Davis (email: Peter.P.Davis@ccmail.census.gov) or
Thomas Mule (email: Vincent.T.Mule.Jr@ccmail.census.gov).

II   Assumptions

Block cluster formation will take place with these underlying assumptions:

A.   The 2000 collection blocks will be clustered.

B.   The waves 1, 2, and 3 of Address Listing field operation are completed and will be used. Wave 4 will be done concurrently with Block Canvassing and will not be used. The housing unit (HU) information has been included on the Master Address File (MAF), but the resulting suffixing may not be done.

C.   No Block Canvassing results will be used for clustering. Wave 1 of Block Canvassing will not be complete by the start of block clustering.

D.   The type of enumeration area (TEA) conversion will not be totally completed.

      1.   Most of the Mailout/Mailback to Update/Leave conversion areas will be identified but there will be no current HU counts available. The current HU counts for these areas will not be determined until after Wave 4 of the Address Listing operation is completed. For these identified areas, the 1990 counts will be available. For some areas, this conversion will not happen until after clustering.

      2.   The Update/Leave to Mailout/Mailback conversion has been delayed. Clustering will not have any of these conversions.

      3.   The Urban Update/Leave conversion is happening now. Additional conversions may occur in subsequent months.

E.   Suffixes will not be used to identify individual blocks. For example, blocks 10001A and 10001B will be treated as block 10001 and assigned to the same block cluster.

F.   Commercial units and Group Quarters will be properly identified as such and not treated as housing units. Commercial units can be identified on the MAF by the variable RESSTAT. Group Quarters can be identified on the MAF by the variables GQHUFLAG, GQ_NAME, MAFSRC, and GQID.

G.   Small block clusters have 0, 1, or 2 HUs. Medium block clusters have between 3 and 79 HUs. Large block clusters have over 79 HUs.

H.   There is insufficient information to identify Crews on Vessels areas to exclude them from the block clustering process.

I.   The most recent Delivery Sequence File (DSF) update was September, 1998. The next DSF update is scheduled for April, 1999.

J.   "Inside the blue line" refers to Mailout/Mailback TEA areas, that is, areas with predominantly city-style addresses. "Outside the blue line" refers to the remaining TEAs, areas with predominantly non-city-style addresses.

K.   The Decennial Master Address File (DMAF) extraction criteria used to define MAF housing unit counts is the ideal in the block clustering process. However, the DMAF extraction criteria have not been determined at the time of clustering. MAF housing unit criteria is specified based on a possible set of rules being considered.

L.   Input files, systems, and results are benchmarked at the time of clustering. Future changes to input files and systems are expected, so discrepancies could occur if comparing clustering results.

M.   A complex algorithm, which is assumed correct, is implemented to identify military blocks. This algorithm gathers information from several sources within TIGER to properly identify military blocks.

N.   During the actual process of forming block clusters, blocks are considered neighbors if they share a line segment boundary. Two blocks joining at a point are not neighbors.

O. If two (2) or more kinds of boundaries separate adjacent blocks, then the following order of boundaries will have priority:

 1. Boundaries listed in Section IV, Paragraph B, Step 1. These boundaries will never be crossed.
 2. Invisible boundaries as defined in Appendix C. These block clusters will be collapsed together.
 3. Areal water ("two-line water"), limited access highway or ridge line. These boundaries will never be crossed.
 4. Streams ("one-line water") or rail lines.
 5. Any remaining boundary.

III Input Files/Systems

The following files/systems will be used in the Census 2000 clustering process.

A. Most recent MAF: This file should include the updates from waves 1, 2 and 3 of Address Listing operations that were just conducted. The last DSF update was September, 1998. No Census 2000 LUCA actions (adds, deletes or corrections) will be reflected since the verification will not be done by the time of clustering. No Block Canvassing results will be used.

B. TIGER system: The TIGER (Topologically Integrated Geographic Encoding and Referencing) system provides proximity, perimeter, and area information for the Census 2000 collection blocks. TIGER itself is a cartographic database with physical features (such as roads, railroads, and rivers) and address ranges. TIGER also contains 1990 housing unit estimates.

IV Block Cluster Specifications

A. Housing Unit Counts

 1. Defining/Tallying MAF Housing Unit Counts

  MAF housing unit counts are required when creating block clusters. The official definition of a MAF housing unit has not yet been finalized. The MAF contains many different types of addresses and address sources. The variable UNITSTAT from the MAF represents the characteristic of an address used in identifying housing units for the purposes of block clustering. Commercial Units and Group Quarters will be excluded using the MAF variables RESSTAT, GQHUFLAG, GQ_NAME, MAFSRC, and GQID.

4

The UNITSTAT codes that will be recognized as valid housing units for block clustering are as follows:

1 = Valid Living Quarters
6 = Under Construction
8 = Vacant Trailer Pad
10 = Boarded Up
11 = Unable to locate
12 = Seasonal
13 = Vacant
15 = Map Spotted Unit with insufficient information*

* A unit with insufficient information includes those addresses outside the blue line that are missing map spot numbers or that have neither a mailing address nor a location description. Include as a valid housing unit those units that have either 1) a mailing address, or 2) a map spot number and a location description.

A complete list of all MAF housing unit codes can be found in Appendix H.

Along with UNITSTAT, field operations will also be used to identify MAF housing units for block clustering as follows:

- For Address Listing areas, an address with one of the UNITSTAT codes listed above will be tallied as a housing unit for block clustering if the address was field verified during Address Listing.

- For areas outside Address Listing, including the three Dress Rehearsal sites, an address with one of the UNITSTAT codes listed above will be tallied as a housing unit for block clustering, including:

    1. addresses whose sole source is the 1990 Address Control File (ACF),
    2. addresses that appear in both the ACF and DSF, and also appear in the most recent DSF,
    3. addresses that appear in both the ACF and DSF, but do not appear in the most recent DSF,
    4. addresses whose sole source is the DSF, and also appear in the most recent DSF,
    5. addresses whose sole source is the DSF, but which do not appear in the most recent DSF.

In addition, for the three Dress Rehearsal sites, include addresses and address actions resulting from the following field operations:

1. Be Counted/Telephone Questionnaire Assistance (TQA)
2. Nonresponse Follow-Up (NRFU)
3. Local Update Census Address (LUCA) Field Verification
4. Targeted Canvass (inside the blue line)
5. Targeted Multi-unit Check (inside the blue line)
6. Update/Leave (outside the blue line)
7. U.S. Postal Service Casing Check

This definition of a MAF HU is intended to be as similar as possible to the final DMAF extraction criteria. However, since the final definition is not known, differences between these two definitions may occur.

2. Determining Housing Unit Counts for Clustering

For block clustering, the most up-to-date counts available will be used. In the past year, Census Bureau personnel performed waves 1, 2 and 3 of the Address Listing operation in most areas outside the blue line; that is, in areas with predominantly non-city-style addresses. The resulting HU counts are considered the most up-to-date. List/Enumerate areas and blocks converted from Mailout/Mailback to Update/Leave were not included in waves 1, 2 or 3 of the Address Listing operation. The only available count for these areas is the 1990 housing unit count. For the remaining blocks, the benefit of Block Canvassing will not be available. No Census 2000 LUCA actions (adds, deletes or corrections) will be used in the block clustering.

Obtain the HU counts to be used for clustering as follows:

a. For blocks included in waves 1, 2 or 3 of the Address Listing operation which include Update/Leave blocks, Rural Update/Enumerate and Update/Leave to Mailout/Mailback conversion blocks: use the housing unit counts from the most recent MAF. These counts were determined in the past year by Census Bureau personnel and are considered the most up-to-date housing unit counts.

b. For List/Enumerate blocks and Mailout/Mailback to Update/Leave conversion blocks: use the 1990 HU counts. In these areas, no other counts are available.

c.   For all remaining blocks which include Mailout/Mailback blocks, Urban Update/Leave blocks, and Military blocks: use the correspondence between 1990 tabulation blocks and 2000 collection blocks to determine which HU count to use as follows:

     i.   One-to-one or a many-to-one correspondence between 1990 tabulation block and 2000 collection block: use the higher of the most recent MAF or 1990 count.

     ii.   One-to-many correspondence between 1990 tabulation block and 2000 collection block: use the most recent MAF count. For the one-to-many correspondence, the 1990 counts in TIGER are estimated to the 2000 blocks based on land area. The most recent MAF gives a more up-to-date number in this instance.

These rules mean that it is possible that an individual block cluster may contain a combination of 1990 and MAF counts. For a summary of HU allocation based on Type of Enumeration Area, see Appendix I.

3.   Tracking Housing Unit Counts

Keeping track of HU counts may provide information that could explain any discrepancies during the sampling process.

a.   Place both the most recent MAF HU counts and the 1990 HU counts on the block cluster output file. The layout for the file is in Appendix A.

For both fields, assign the following:

     If the source has a HU count, assign the HU count (0-99999)
     Else if the source does not have a count, assign a blank ('   ').

Note: An example of when a source does not have a HU count is List/Enumerate area where there is no MAF count.

7

b.   Place an indicator variable for each block on the output file. It will indicate if the housing unit count used for clustering comes from the most recent MAF or the 1990 counts.

For this field, assign the following:

> If the count is from the most recent MAF, assign the Housing Unit Count Indicator = '1'
> Else if the count is from the 1990 count, assign the Indicator = '2'.

B.   Formation Rules

This formation is hierarchical. That is, step 1 takes precedence over step 2 and so on.

1.   Block clustering will adhere to several geographical constraints. Block clusters will not cross the following boundaries:

a.   County and sub-county partitions in the TIGER system: counties will be a boundary. For large counties, The TIGER system maintains county parts on separate files. These county parts are referred to as sub-county partitions. In these counties, a block cluster will not cross the sub-county partition. A list of these counties is given in Appendix B.

b.   Census Tract: PES block clusters will respect the 6 digit (including suffixes) tract boundaries for 2000. The tract definitions GEO will use to cluster are the interim tract boundaries. Interim tracts are the 1990 tracts adjusted for the 2000 collection block boundaries and are also referred to as pseudo-tracts. Census 2000 tracts will not be defined until some time in 2000.

c. Groups of Type of Enumeration Areas (TEA): The following TEAs may be clustered together.

TEA Group	Description	TEA
A	Mailout/Mailback	1
	Urban Update/Leave	7
	Update/Leave to Mailout/Mailback conversions	8
B	Update/Leave	2
	Rural Update/Enumerate	5
C	List/Enumerate	3
D	Military (outside the blueline)	6
E	Mailout/Mailback to Update/Leave conversions	9

d. Military Areas: military blocks should be clustered with military blocks. For military areas classified as Military Type of Enumeration Area, this is accomplished by the TEA restriction, above. A complex algorithm, assumed to be correct, gathers a set of TIGER information used to identify military blocks inside the blue line.

e. American Indian Country (AIC) Land: Block clusters on AIC land will respect the American Indian Reservation/Alaska Native Village Statistical Area (ANVSA) values from TIGER. The values are as follows:

AIR/ANVSA Values	Element Description
Blank	No American Indian Country specified
0001-4989	American Indian reservation/trust land
5001-5989	Tribal jurisdiction statistical area
6001-5989	Alaska Native Village Statistical Area
9001-9589	Tribal designated statistical area

    i. American Indian Reservation (AIR) blocks are allowed to be clustered with other AIR blocks on the same reservation.

    ii. Blocks on AIC but outside AIR are allowed to be clustered with other AIC but outside AIR blocks.

    iii. The blocks must have the same ANVSA value to be combined.

iv.   If a block is partially on an American Indian Reservation, classify the block as AIR. If the block is partially on an American Indian Country and not in an AIR then classify as AIC.

v.   Blocks that contain two AIRs are assigned only one AIR value. The rule for designating these blocks to only one AIR value is accomplished in a complex algorithm designed by the TIGER Systems Branch and is assumed correct.

2.   Exclude the following blocks from the clustering procedure. They will not be clustered with any adjacent blocks, and they will not be included on the verification and the block cluster output files sent to Decennial Systems and Contracts Management Office (DSCMO).

    a.   Blocks consisting entirely of bodies of water (water blocks)
    b.   Blocks in Remote Alaska [TEA=4]

3.   All blocks within each county separated by invisible boundaries will be collapsed except when the invisible boundary is a block extension identified by Census feature codes F20 (Feature Extension–extensions not otherwise classified), F21 (Automated extension), or F22 (Irregular block extension). All Feature Class F codes besides F20, F21 and F22 will be collapsed. Intermittent streams (Feature Code H12) and Intermittent canals, ditches, or aqueducts (Feature Code H22) forming block boundaries will also be considered to be invisible boundaries. See Appendix C for the invisible boundary codes.

4.   All blocks with 80 or more housing units will be block clusters by themselves. No other blocks will be clustered with this block unless the block completely surrounds another block with zero HUs. (see rule #9.)

5.   Any block larger than 15 square miles will be a block cluster by itself unless the block cluster completely surrounds another block. (see rule #9.)

6.   Block clusters that are not contiguous with any other block clusters within the boundaries specified in step 1 will be block clusters by themselves.

7.   Follow these guidelines in hierarchical order when combining neighboring blocks:

    a.   Never cross areal water ("two-line water"), limited access highways, and ridge lines.
    b.   Do not create a block cluster with 80 or more housing units.
    c.   Do not create a block cluster with more than 15 square miles. (Exception: see Rule #9.)

d.   Never cross streams ("one-line water") or rail lines except when there are no other adjacent clusters available for clustering. If a cluster must cross either a stream or a rail line, it should cross the stream before the rail line.

8.   Sometimes, a block can completely surround another block. The block which is enclosed within the surrounding block can be identified because it has only one neighbor. If a block has only one neighbor, combine as follows:

     a.   If the surrounding block has fewer than 80 housing units (even blocks with 0 housing units), then collapse the surrounding block with the enclosed block as long as the resulting block cluster does not exceed 80 HUs.

     b.   If the surrounding block has 80 or more housing units, then:

         i.   collapse the surrounding block cluster with the enclosed block only if the enclosed block has a total of 0 housing units.

         ii.   otherwise, do not combine.

9.   All blocks with more than 15 square miles:

     a.   are NOT eligible to initiate the clustering algorithm.
     b.   are NOT considered eligible neighbors for clustering.
     c.   are NOT eligible to be combined when they are surrounded by another block.
     d.   are eligible to be combined with a block when it surrounds another block (see rule #9).
     e.   are NOT eligible to absorb block clusters with 0 housing units along the perimeter.

10.   All blocks containing 3 to 25 housing units:

     a.   will initiate the clustering algorithm.
     b.   are eligible to be collapsed with neighboring blocks such that the resulting cluster does not exceed 80 HUs.
     c.   are eligible to be combined when they are surrounded by another block (see rule #9).
     d.   are eligible to be combined with a block when it surrounds another block (see rule #9).
     e.   are eligible to absorb block clusters with 0 housing units along the perimeter.

11.　　All blocks containing 26 to 79 housing units:

    a.　　do NOT initiate clustering.
    b.　　are NOT considered eligible neighbors for clustering.
    c.　　are eligible to be combined when they are surrounded by another block ·
        (see rule #9).
    d.　　are eligible to be combined with a block when it surrounds another block
        (see rule #9).
    e.　　are allowed to absorb blocks with 0 housing units along the perimeter.

12.　　All blocks with 1 or 2 housing units:

    a.　　do NOT initiate the clustering algorithm.
    b.　　are eligible for collapsing with neighboring block clusters only if the
        adjacent block cluster contains at least 3 housing units.
    c.　　are eligible to be combined when they are surrounded by another block
        (see rule #9).
    d.　　are eligible to be combined with a block when it surrounds another block
        (see rule #9).
    e.　　are allowed to absorb blocks with 0 housing units along the perimeter.

13.　　All blocks with 0 housing units:

    a.　　are NOT eligible to initiate the clustering algorithm.
    b.　　are NOT eligible to be a neighboring block cluster.
    c.　　are eligible to be combined when they are surrounded by another block
        (see rule #9).
    d.　　are eligible to be combined with a block when it surrounds another block
        (see rule #9).
    e.　　are NOT eligible to absorb block clusters with 0 housing units along the
        perimeter.

14.　　All blocks with 80 or more HUs:

    a.　　are NOT eligible to initiate the clustering algorithm.
    b.　　are NOT eligible to be a neighboring block cluster.
    c.　　are NOT eligible to be combined when they are surrounded by another
        block (see rule #9).
    d.　　are eligible to be combined with a block when it surrounds another block
        (see rule #9).
    e.　　are NOT eligible to absorb block clusters with 0 housing units along the
        perimeter.

Note:　See Appendix E for table summaries of Formation Rules 8 through 14.

C.   Algorithm

The goal of the algorithm is to cluster blocks and generate clusters that have on average 30 HUs per block cluster for block clusters with 3 to 79 housing units. First, blocks are prepared and screened to identify which phases of processing the block will undergo. There are four basic phases of processing: 1) calculating target size, 2) clustering, 3) checking for enclosed blocks, and 4) a zero block perimeter search. The algorithm proceeds as follows.

1.   Calculate the Clustering Target for the county

The overall goal of clustering is to have an average of 30 HUs per cluster for the medium clusters. Since many blocks have more than 30 HUs, if clusters of size 30 are formed, the overall average will be greater than 30. Therefore, a clustering target for blocks containing 3 to 25 HUs which balances the number of blocks greater than or equal to 26[1] is determined. Do this as follows:

a.   For each county, classify blocks that have between 3 and 79 HUs into two groups: 1) between 3 and 25 HUs inclusive and 2) between 26 and 79 HUs inclusive. For each group, count the number of blocks and the number of HUs. Blocks that are between 26 and 79 HUs will be considered block clusters by themselves. This can be used to determine at what average does the algorithm need to cluster the 3 to 25 HU blocks so that the overall medium block cluster average is as close to 30 as possible.

Calculate the Clustering Target:

$$\text{Cluster Target} = \frac{(\# \text{ of } HU_{3\text{-}25})}{\left[\left(\dfrac{\# \text{ of } HU_{3\text{-}25} + \# \text{ of } HU_{26\text{-}79}}{30}\right) - \# \text{ of blocks}_{26\text{-}79}\right]}$$

If the number of blocks containing 3 to 25 HUs is zero, then proceed to Step 2, Early Stages of Block Clustering. If the Cluster Target is less than 10, set the Cluster Target to 10. If the sum of the number of HUs between 3 to 25 plus the number of HUs between 26 to 79, divided by 30, equals the number of blocks containing 26 to 79 HUs, then set the Cluster Target to 26.

---

[1] 26 is used instead of 30 because 26 is within 15% of 30. 15% is considered sufficiently close to the target to stop clustering.

Note: This average is calculated before the invisible boundaries are collapsed.

b. For example, a county has 5479 blocks that have between 3 and 25 HUs per block. In these 5479 blocks, there are 77,696 HUs. There are 4087 · blocks that have between 26 and 79 HUs per block. In these 4087 blocks, there are 167,103 HUs.

The cluster target would be calculated as follows:

$$\text{Cluster Target} = \frac{77696}{\left[\left(\dfrac{77696 + 167103}{30}\right) - 4087\right]} = 19.1$$

The 3 to 25 HU blocks would be clustered with a desired average of 19.1 HUs per block cluster. This will produce a medium block cluster average that is close to the goal of 30 HUs per block cluster.

c. This average is calculated separately for each county processed. Since counties can have diverse block HU density distributions, the medium block cluster average goal of 30 HUs is better achieved by having each county's medium block clusters average around 30.

d. Blocks with 1 or 2 HUs are eligible to be combined if they are the closest neighbor. These HUs were not used in computing the above average because 1) the number of 1 or 2 HU blocks to be combined is unknown and 2) the total number of HUs in blocks with 1 or 2 HUs is very small as compared to the total number of HUs in the medium blocks.

2. Early stages of block clustering: Prepare and Screen the Blocks

a. The algorithm checks the geographic constraints first. These boundaries are never crossed. The next step is to collapse the blocks that are separated by invisible boundaries.

b. The 80+ HU blocks and the 15 sq. mile blocks are separated and considered clusters themselves, not eligible to initiate clustering. Block clusters containing either 80+ HUs are eligible to be collapsed if they surround a block containing zero HUs. Blocks spanning more than 15 sq. miles are eligible to be collapsed only if they completely surround a block and meet the housing unit requirements (Formation Rule #9). No other block clusters will be clustered with either an 80+ HU block or a 15 sq mile block.

14

c. Blocks between 26 and 79 HUs are considered to be clusters by themselves, not eligible to initiate clustering. These block clusters are eligible to be collapsed if they 1) completely surround another block or another block completely surrounds it and 2) satisfy the enclosure rules (Formation Rule #9). These blocks are eligible to perform the zero neighbor perimeter search.

d. The small blocks (0, 1, or 2 HUs) are set aside for the time being as individual clusters. They will be eligible to be clustered later on. Any that are not collapsed remain as small block clusters.

3. Form block clusters:

a. This stage begins with the blocks containing 3 - 25 HUs in the TIGER order. All blocks with the number of housing units greater than 85% of the cluster target, calculated in part 1, are considered to be sufficiently close to the preferred average block cluster HU size and although they do not initiate clustering, they are eligible neighbors for collapsing.

b. Given a block, call it block A, that has fewer HUs than 85% of the cluster target, identify an eligible list of neighbors. Collapse block A with its closest neighbor. The closest neighbor is defined by the block with the closest centroid that has at least 1 but no more than 25 HUs and a shared line segment boundary.

c. Let block A's closest neighbor be block B. A new block cluster, call it AB, is formed by the combination of Block A and its neighboring block, block B. If the total number of housing is greater than 85% of the cluster target then proceed to Step 4: Enclosed Blocks. If the total number of housing units is less than 85% of the cluster target then find the closest neighbor, block C, and form ABC. Continue to find the closest neighbor and collapse it into the block cluster until the total number of housing units is greater than 85% of the cluster target.

d. Once a block cluster reaches 85% of the cluster target, it is written off to the used file and removed from the neighbor list. This block cluster will not be an eligible neighbor for any ensuing blocks in the TIGER order. The newly formed block cluster will then proceed to the enclosed block and perimeter search steps.

4. Collapse Enclosed Blocks:

    a. Definition: Enclosed blocks are blocks with only one neighbor.

    b. Collapse the enclosed block with its surrounding block if the enclosed block has < 80 HUs and the resulting cluster is not > 80 HUs. If the enclosed block has zero units, it can be combined with an 80+ HU block.

5. Zero Neighbor Perimeter Search:

Once a block cluster has been collapsed with its closest neighbors to contain more than 85% of the cluster target and checked for enclosed blocks, the final step involves searching the perimeter of the block cluster. If there is a neighboring block on the perimeter that has 0 housing units, then the 0 housing unit block is to be collapsed into the block cluster.

6. Proceed to the next block that has fewer HUs than 85% of the cluster target, and restart the clustering process at step 3b, above.

V   Identification

A.   Block Cluster Number

Within county, each block cluster will be uniquely identified based on a numbering process which uses the cluster's latitude and longitude. The latitude and longitude are merged, sorted, and then transformed into a 5-digit GEO cluster number. This cluster number will produce a geography sort.

B.   Collapsing Across Invisible Boundaries

Block clusters that were formed by collapsing invisible boundaries need to be identified (see Appendix C.) A block cluster may have 1) more than 1 block and 2) more than 80 HUs or 15 square miles. One reason for this occurrence is if invisible boundaries are collapsed. The indicator variable will allow us to verify this.

Assign the Invisible Boundary Indicator field to the output file:

1. If an invisible boundary is crossed, assign a value of '1'.
2. If no invisible boundaries are crossed, assign a value of '0'.

C. American Indian Country Block Clusters

Block clusters that contain American Indian Country need to be identified. These are lands that are American Indian Reservation/trust land, tribal jurisdiction statistical area, tribal designated statistical area, and Alaska native village statistical area. American Indians will have their own sampling stratum in 2000 and hence need their own identification for the clustering process.

Use the variable American Indian Reservation/Alaska Native Village Statistical Area (ANVSA) to identify the types of American Indian Country. This variable is defined in TIGER Documentation: Chapter III, Section B, TIGER System County Partition Data Element Definitions. Assign the American Indian Country field to the output file as follows:

AIC Indicator	ANVSA Values	Element Description
0		No American Indian Country specified
1	0001-4989	American Indian reservation/trust land
2	5001-5989	Tribal jurisdiction statistical area
	6001-5989	Alaska Native Village Statistical Area
	9001-9589	Tribal designated statistical area

D. Type of Enumeration Area

Put two Type of Enumeration Area codes on the cluster output file. The first is the Initial Block TEA [values: 1, 2, 3, 4, 5, 6, 7, 8 or 9]. This is the block TEA value at the time clustering occurred. The second is the TEA Group variable [values: A, B, C, D, or E (See Formation Rule 1c)]. This is the TEA group value that is assigned at the time of clustering.

17

E.   Military Area Indicator

Blocks on Military Areas need to be identified. Military areas are a boundary in step 1 of the Formation Rules. Put an indicator on the output file to denote if the block is a military area. This provides a check that the rule was implemented correctly. Outside the blue line, military areas are designated by TEA value of 6. For military blocks inside the blue line, Geography division will use TIGER to identify them.

Assign the Military Area Indicator to the file as follows:

1.   If no military area in block, then assign value of '0'.
2.   If block contains military area then assign value of '1'.

VI   Output

A.   Equivalency Files

A correspondence between 1990 Tabulation and 2000 collection blocks is required so that DSCMO can determine the approximate demographic composition of each block cluster. Make available the standard Block Equivalency File in the standard format relating the 2000 collection blocks to the 1990 tabulation blocks. The format for the equivalency file is located in Appendix J.

B.   Verification Maps

GEO will deliver block cluster maps for review by the Decennial Statistical Studies Division (DSSD) for use in verifying the block clustering for Census 2000. There will be one multi-page map for each county requested. (See section D., Testing and Production, below.) At a minimum, these maps should include block boundaries, block numbers, block cluster boundaries, and block cluster numbers. Also included on the maps will be the following color designations for these respective Type of Enumeration Areas: TEA 1, TEA 7, and TEA 8 should be similar shades of blue, TEA 2 and TEA 5 should be similar shades of red, TEA 3 should be yellow, TEA 6 should be green, and TEA 9 should be purple.

C.   Block Files

The block clustering operation will be a flowing process involving GEO, DSSD, and DSCMO. The verification process will be based upon reviewing maps, verification files and summary statistics. Once the clustering file results have been reviewed and approved for a state then the completed files will be provided from GEO to DSCMO. DSSD will give official approval for GEO to make files available to DSCMO.

18

1.   Verification Files

The files delivered by GEO to DSSD for Census 2000 will be similar to that of the cluster files for the 1998 Dress Rehearsal. Changes have been made to add the indicator fields created and the housing unit count information.

A selected number of counties for each state will be specified for review by DSSD. Each file will contain one record for each block for the selected counties in the state. (See section D., Testing and Production.)

2.   File to DSCMO

After a state has been verified, GEO will deliver the complete file (all blocks in that state) to DSCMO staff. The file should contain all of the fields listed in the block cluster verification file layout. (See Appendix A.)

Place the file in the GEBA01::GEO_PUBLIC:[PES.BC] subdirectory. DSCMO will be able to obtain the file from there.

D.   Testing and Production

For testing purposes, several counties from the initial stages of block clustering will be reviewed. This testing procedure will examine these counties to check the rules, the algorithm, and the overall goal of block clustering. This procedure will occur prior to production.

DSSD will use the maps and GusX to visually inspect the block clusters. Using the verification output file, a SAS program will be written to identify any discrepancies in the rules, algorithm, and/or goal of block clustering. The TIGER Systems Branch will generate a Cluster Summary File that contains the number of HUs, the number of blocks, the number of zero HU blocks, the Cluster Target, the number of HUs between 3 and 25, the number HUs between 26 and 79, and the number of blocks between 26 and 79 for each cluster.

For testing, the Dress Rehearsal sites will be run and made available for review. Also, DSSD will examine one county from each of the Wave 1 states. The counties that will be tested are as follows: Washington, DC, Hennepin County, MN, Denali Borough, AK, Glacier, MT, Mountrail, ND, Shannon, SD, Florence, WI, and Washington, ID. For two counties, GEO will produce maps. For the remaining counties, DSSD staff can visit GEO and use GusX to review these counties instead of producing maps.

During production, DSSD will examine two counties from each of the Wave 1 states. The verification and Cluster Summary File for each county and the state summary file will be sent to DSSD after the state is processed. The Cluster Summary File only needs to be produced for the production check counties. One county from each state is examined during testing. If DSSD compares the testing to the production files in these counties, no differences should exist. Hence, no maps for these counties are necessary. Maps will only be needed for the other counties.

DSSD will examine one county from each of the Wave 2 states. Maps, verification file, the Cluster Summary File and the state summary file will be sent to DSSD.

For Wave 3 and 4 states, one county in California, Illinois and Texas will be completely reviewed. Maps, the county verification file, the Cluster Summary File and the state summary file will be sent to DSSD. For the remaining states and Puerto Rico, one county has been chosen to be computer checked. The county verification file, Cluster Summary File and state summary file will be sent to DSSD. No maps need to be made for these counties.

Appendix F contains the order in which the states will be processed during testing and production. Appendix G contains the counties to be verified during production.

E.   Summary Counts

GEO will produce summary block cluster counts for DSSD to use in monitoring future sampling operations. Also, it will allow DSSD to examine summary results of the clustering in every county/partition in the state during verification. Having these at the county/partition level will allow DSSD to combine them as needed. The Cluster Target calculated for each county/partition will also be allocated to this file. This will document the cluster target used for each county/partition.

This file will be an ASCII file for each state with one record per partition and will be sent with the verification materials. Unlike the verification files, all counties/partitions will be listed on this file. The layout of the file is in Appendix D.

VII   After Sampling File Preparation

GEO will send the block cluster files to DSCMO, which will then select a sample of clusters. After sampling, DSCMO will send a file of the sampled clusters to GEO so that three fields can be added to the file. GEO will assign 1) revised block TEA, 2) revised city-style address indicator and 3) the Local Census Office (LCO) code. This information will be used in the creation of the Collection Geographic Reference File (GRF). Return the updated sample file to DSCMO.

A. Revised Block TEA

The block TEA values may change after the block clustering process is completed. After the sample is selected, the TEA codes for the sampled block clusters need to be updated.

Assign the Revised Block TEA (RBTEA) value for each block.

B. Cluster TEA Code

PES operations handle block clusters that have city-style addresses differently than block clusters with non-city-style addresses. Because of this, a Cluster TEA code is needed to identify city-style address clusters and non-city-style address clusters. Clusters that have at least 1 block that is Update/Leave (RBTEA=2), List/Enumerate (RBTEA=3), Rural Update/Enumerate (RBTEA=5), or Block Canvassing moved to Address Listing (RBTEA=9) are considered to be non-city-style address clusters. Clusters that contain solely blocks that are Mailout/Mailback (RBTEA=1, 6 or 8) or Urban Update/Leave (RBTEA=7) are considered to be city-style address clusters. This is a cluster-level variable. Each block in the cluster will receive the same cluster value. Assign the city-style address indicator to each block as follows.

If a cluster has at least one block with an RBTEA = 1, 7 or 8 then assign the Cluster TEA Code = '2' (non-city-style address) to all blocks in the cluster; otherwise, assign the Cluster TEA Code = '1' to all blocks in the cluster.

C. Local Census Office (LCO) Code

After the PES sample is selected, the Census LCO field will need to be updated. For each block on the file, assign the LCO code to the file.

cc: DSSD Census 2000 Memorandum Series Distribution List
PES Implementation Team
Statistical Design Team Leaders
Sample Design Team
C. Hantman   (GEO)
R. Ruiz   (GEO)
S. Holt   (GEO)
K. Todd   (GEO)

Appendix A
Verification and DSCMO Block Cluster File Layout

The following is the layout of the block clustering file that will be made available to DSCMO. The verification files will include all of the records for the counties specified by DSSD. There will be a record for each block. When a state is verified, all of the county files will be combined into one file and made available to DSCMO.

Variable	Location
State	1:2
County	3:5
Interim Tract (a.k.a. pseudo-tract)	6:11
Block Number	12:16
Blank	17:17
Cluster Number	18:22
Blank	23:23
Cluster Size code	24:24

      1 = Clusters with 0 HUs
      2 = Clusters with 1 HUs
      3 = Clusters with 2 HUs
      4 = Clusters with between 3 and 5 HUs
      5 = Clusters with between 6 and 9 HUs
      6 = Clusters with between 10 and 19 HUs
      7 = Clusters with between 20 and 29 HUs
      8 = Clusters with between 30 and 79 HUs
      9 = Clusters with 80 or more HUs

Variable	Location
Blank	25:25
Block Area (Sq. Miles)	26:31
Blank	32:32
Block Perimeter (Miles)	33:37
Blank	38:38
Block Cluster Area (Sq. Miles)	39:44
Blank	45:45
Block Cluster Perimeter (Miles)	46:50
Number of HUs in cluster	51:55
Number of HUs in block	56:60

Variable	Location
Block TEA	61:61

    1 = Mailout/Mailback
    2 = Update/Leave
    3 = List/Enumerate
    5 = Rural Update/Enumerate
    6 = Military
    7 = Urban Update/Leave
    8 = Update/Leave to Mailout/Mailback conversions
    9 = Mailout/Mailback to Update/Leave conversions

TEA Group for Block Cluster   62:62
    A= Mailout/Mailback or
        Update/Leave to Mailout/Mailback conversions
    B= Update/Leave or
        Rural Update/Enumerate
    C=List/Enumerate
    D=Military
    E=Mailout/Mailback to Update/Leave conversions

2000 MAF HUs count   63:67
    ' ' Blank if no HU count available

1990 ACF HUs count   68:72
    ' ' Blank if no HUs count available

Housing Unit Count Indicator   73:73
    1 = from 2000 MAF
    2 = from 1990 ACF

Invisible Boundary Collapse Indicator   74:74
    0 = No
    1 = Yes (Collapsing across Invisible Boundary in BC)

American Indian Country Indicator   75:75
    0 = No American Indian Country
    1 = American Indian Reservation/trust land
    2 = Tribal jurisdiction statistical area/
        Alaska Native Village statistical area/
        tribal designated statistical area

Military Indicator   76:76
    0 = No Military Area
    1 = Block contains Military Area

23

Appendix B
Partitioned Counties in the TIGER system

The following counties are partitioned into smaller sub-county geographic areas in the TIGER system. The block clustering in these counties will have to observe these sub-county partition boundaries.

04013 Maricopa, AZ
04019 Pima, AZ
06029 Kern, CA
06037 Los Angeles, CA
06059 Orange, CA
06065 Riverside, CA
06071 San Bernardino, CA
06073 San Diego, CA
06085 Santa Clara, CA
12025 Dade, FL
12099 Palm Beach, FL
15003 Honolulu, HI
17031 Cook, IL
17043 DuPage, IL
25017 Middlesex, MA
26163 Wayne, MI
36059 Nassau, NY
36103 Suffolk, NY
39049 Franklin, OH
42003 Allegheny, PA
48029 Bexar, TX
48113 Dallas, TX
48157 Fort Bend, TX
48201 Harris, TX
48439 Tarrant, TX
53033 King, WA

Appendix C
Block Cluster Invisible Boundary Codes

The following is the list of boundary codes that are to be treated as invisible boundaries during the clustering process.

- F00 - Nonvisible boundary, classification unknown or not elsewhere classified
- F10 - Nonvisible governmental unit boundary
- F11 - Offset corporate boundary
- F12 - Corporate corridor
- F13 - Nonvisible interpolated boundary
- F14 - Superseded political boundary
- F15 - Corrected governmental unit boundary
- F16 - EAC nonvisible boundary
- F17 - State legislative non-visible boundary
- F18 - Congressional District non-visible boundary
- F23 - Closure extension
- F24 - Nonvisible separation line
- F25 - Nonvisible corporate corridor centerline
- F30 - Point-to-point line
- F40 - Property line
- F50 - ZIP Code boundary
- F60 - Map edge
- F70 - Statistical boundary
- F71 - 1980 statistical boundary
- F72 - 1990 block boundary
- F73 - Urbanized area land use boundary
- F74 - 1990 Statistical Boundary
- F80 - Nonvisible other tabulation boundary, major category used when the minor category could not be determined
- F81 - School district boundary
- F82 - Special census tabulation boundary
- F83 - Census 2000 Collection Block Boundary
- F84 - Census 2000 Statistical Area Boundary
- F85 - Census 2000 Tabulation Block Boundary
- F86 - Local Administrative Line

- H12 - Intermittent streams or wash
- H22 - Intermittent canal, ditch, or aqueduct

Appendix D
Summary File Layout

The following is the layout for the partition-level summary statistics requested for each state. There will be one record for each partition. In most states, the partition is equivalent to the county. However, in some states, a county has been subdivided into partitions. (See Appendix B. for counties that are partitioned.) Hence, a partition-level summary is necessary.

Variable	Location
Partition	1:1
State (FIPS code)	3:4
County (FIPS code)	6:8
Number of Medium BCs (Non-Indian)	10:15
Number of Large BCs (Non-Indian)	17:22
Number of Small BCs (Non-Indian)	24:29
Number of Medium BCs with AIR but no other Indian Country land	31:36
Number of Medium BCs with some Indian Country but no AIR land	38:43
Number of Large BCs with AIR but no other Indian Country land	45:50
Number of Large BCs with some Indian Country land but no AIR land	52:57
Number of Small BCs with AIR but no other Indian Country land	59:64
Number of Small BCs with some Indian Country but no AIR land	66:71
Number of 'Water' and Remote Alaska Blocks removed	73:78
Number of HUs on Medium BCs (Contain No American Indian Land)	80:85
Number of HUs on Large BCs (Contain No American Indian Land)	87:92
Number of HUs on Small BCs (Contain No American Indian Land)	94:99
Number of HUs on Medium BCs on American Indian Reservations only	101:106
Number of HUs on Large BCs on American Indian Reservations only	108:113
Number of HUs on Small BCs on American Indian Reservation only	115:120
Number of HUs on Medium BCs on American Indian Country land but not an American Indian Reservation.	122:127
Number of HUs on Large BCs on American Indian Country land but not an American Indian Reservation.	129:134
Number of HUs on Small BCs on American Indian Country land but not an American Indian Reservation	136:141
Number of HUs on "water" or Remote Alaska blocks	143:148
Average Number of HUs per BC for BCs with 3 or more HUs	150:155
Average Number of HUs per BC for BCs with 3 to 79 HUs	157:162
Cluster Target	164:169

Summary of Formation Rules 8 - 14

Block Area (Sq. Miles)	Initiate Clustering?	Eligible to collapse with neighbors?	Eligible to be absorbed if completely surrounded ?	Eligible to check if completely surrounds a block?	Absorb zero blocks along perimeter?
15 +	No	No	No	Yes[1]	No
Less than 15	Follow HU Rules Below				

[1] Not to exceed 80 HUs.

Block Size (HUs)	Initiate Clustering?	Eligible to collapse with neighbors?	Eligible to be absorbed if completely surrounded ?	Eligible to check if completely surrounds a block?	Absorb zero blocks along perimeter?
0	No	No	Yes[1]	Yes[1]	No
1	No	Yes[1,3]	Yes[1]	Yes[1]	Yes
2	No	Yes[1,3]	Yes[1]	Yes[1]	Yes
3 - 25	Yes	Yes[1,3]	Yes[1]	Yes[1]	Yes
26 - 79	No	No	Yes[1]	Yes[1]	Yes
80+	No	No	No	Yes[2]	No

[1] Not to exceed 80 HUs.
[2] Collapse only if there are no housing units in the enclosed block.
[3] Not to exceed 15 Square Miles.

Appendix F
The Production Process of States by Wave

For production, states will be processed in the order listed below. It is anticipated review will. ˜ follow a similar order. Wave 1 begins with Alaska as the first state to be verified. Once work on Alaska is completed, Idaho is the next state. The production process continues through Wave 4.

Wave 1	Wave 2	Wave 3	Wave 4
Alaska	Arkansas	Delaware	Puerto Rico
Idaho	Connecticut	Maryland	
Minnesota	Hawaii	Ohio	
Montana	Kentucky	South Carolina	
North Dakota	Louisiana	Alabama	
South Dakota	Massachusetts	Florida	
Washington, DC	Mississippi	Georgia	
Wisconsin	Rhode Island	Illinois	
Wyoming	Tennessee	Indiana	
	West Virginia	Iowa	
	Arizona	Kansas	
	Colorado	Michigan	
	Maine	Missouri	
	Nebraska	New Jersey	
	Nevada	North Carolina	
	New Hampshire	Oklahoma	
	New Mexico	Texas	
	New York	Virginia	
	Oregon	California	
	Pennsylvania		
	Utah		
	Vermont		
	Washington		

Appendix G
Production Counties for Review

Wave 1

Two counties will be reviewed for each state in Wave 1 Processing. A county name in Italics indicates that the county was reviewed during testing and maps do not need to be produced for these counties. During production, only the county verification file and the state summary file needs to be sent to DSSD. The remaining counties will be a complete review with maps, verification files and summary files.

Alaska (FIPS state code 02)
> Wade Hampton Census Area (FIPS county code 270)
> *Denali Borough (068)*

District of Columbia (11)
> *District of Columbia (001)*

Idaho (16)
> Minidoka (068)
> *Washington (087)*

Minnesota (27)
> *Hennepin (053)*
> Watonwan (165)

Montana (30)
> *Glacier (035)*
> Yellowstone (111)

North Dakota (38)
> Morton (059)
> *Mountrail (061)*

South Dakota (46)
> Brown (013)
> *Shannon (113)*

Wisconsin (55)
> *Florence (037)*
> Racine (101)

29

Wave 2

One county will be reviewed for states that are in Wave 2 of the Address Listing operation. This will be a complete review with maps, verification files and summary files.

State	County
Arkansas (05)	Lee (077)
Arizona (04)	Apache (001)
Colorado (08)	Conejos (021)
Connecticut (09)	Windham (015)
Hawaii (15)	Kauai (007)
Kentucky (21)	Union (225)
Louisiana (22)	Madison (065)
Massachusetts (25)	Suffolk (025)
Maine (23)	Piscataquis (021)
Mississippi (28)	Jefferson (063)
Nebraska (31)	Thurston (173)
New Hampshire (33)	Carroll (003)
New Mexico (35)	Guadalupe (019)
New York (36)	Franklin (033)
Nevada (32)	Humboldt (013)
Oregon (41)	Jefferson (031)
Pennsylvania (42)	Forest (053)
Rhode Island (44)	Bristol (001)
Tennessee (47)	Haywood (075)
Utah (49)	San Juan (037)
Vermont (50)	Essex (009)
Washington (53)	Franklin (021)
West Virginia (54)	McDowell (047)
Wyoming (56)	Carbon (007)

Wave 3

The review for Wave 3 will be in two parts. A complete review will be done of one county in California, Illinois and Texas. Maps, verification files and summary files will be generated for these counties. The remaining states will have one county checked by the SAS program. No maps will be generated for these counties. Only the verification and summary files will be sent to DSSD.

Complete Review

State	County
California (06)	San Fransisco (075)
Illinois (17)	Cook (031)
Texas (48)	El Paso (141)

Note: Cook County, IL is partitioned on TIGER. Pick one of the sub-county partitions for the review.

Computer Program Review Only

State	County
Alabama (01)	Macon (087)
Delaware (10)	Kent (001)
Florida (12)	Gadsden (039)
Georgia (13)	Hancock (141)
Iowa (19)	Muscatine (139)
Indiana (18)	Grant (053)
Kansas (20)	Grant (067)
Maryland (24)	Somerset (039)
Michigan (26)	Saginaw (145)
Missouri (29)	Pemiscot (155)
North Carolina (37)	Warren (185)
New Jersey (34)	Salem (033)
Ohio (39)	Erie (043)
Oklahoma (40)	Adair (001)
South Carolina (45)	Allendale (005)
Virginia (51)	Charles City (036)

Wave 4

The Wave 4 review will be a computer program review only.

State	County
Puerto Rico (72)	Florida Municipio

Appendix H
MAF Housing Unit Status Code

This appendix lists the housing unit status codes. The variable UNITSTAT, as allocated on the MAF, is the field that identifies the housing unit status of an address.

UNITSTAT Legal values:

    1 = Valid Living Quarters
    2 = Demolished
    3 = Open to the elements
    4 = Nonexistent
    5 = Provisional Add
    6 = Under Construction
    7 = Duplicate
    8 = Vacant Trailer Pad
    9 = Burned Out
    10 = Boarded Up
    11 = Unable to locate
    12 = Seasonal
    13 = Vacant
    15 = Map Spotted Unit with insufficient information
    30 = MAF unit moved to another MAF partition
    31 = Other uninhabitable

Appendix I
Rules for Housing Unit Count Allocation
During PES Clustering

Type of Enumeration Area	Rules/Conditions	Clustering HU count
TEA 1: Mailout/Mailback	1-to-1 or many-to-1 correspondence between 1990 tab block and 2000 collection block	use higher of MAF or 1990 count
	1-to-many correspondence	use MAF count
TEA 2: Update/Leave		use MAF count
TEA 3: List/Enumerate		use 1990 count
TEA 4: Remote Alaska	exclude from PES	
TEA 5: Rural Update/Enumerate		use MAF count
TEA 6: Military	1-to-1 or many-to-1 correspondence between 1990 tab block and 2000 collection block	use higher of MAF or 1990 count
	1-to-many correspondence	use MAF count
TEA 7: Urban Update/Leave	1-to-1 or many-to-1 correspondence between 1990 tab block and 2000 collection block	use higher of MAF or 1990 count
	1-to-many correspondence	use MAF count
TEA 8: Update/Leave to Mailout/Mailback conversion		use MAF count
TEA 9: Mailout/Mailback to Update/Leave conversion		use 1990 count

Appendix J
Equivalency File Layout

The following is the record layout for the Block Equivalency File relating 1990 tabulation blocks to 2000 collection blocks.

Field	Type	Length	Description
ST	CHAR	2	State Code from GTUBAN
RS1		1	Space
COU	CHAR	3	County code from GTUBAN
RS2		1	Space
TRACTBAS	CHAR	4	Tract/Block Numbering area base from GTUBAN
TRACTSUF	CHAR	2	Tract/Block Numbering area suffix from GTUBAN
RS3		1	Space
BLOCKBAS	CHAR	3	Block Base from BKARA
TAB90SUF	CHAR	1	Block Suffix from BKARA
RS4		1	Space
ST2	CHAR	2	State code from COL2000
RS5		1	Space
COU2	CHAR	3	County code from COL2000
RS6		1	Space
COBLKBAS	CHAR	5	2000 Collection Block base from COL2000